Business Roles in the Emerging Open-Data Ecosystem

Juho Lindman is an assistant professor of information systems science at the Hanken School of Economics. His research focuses on open source software development, open data, open access, and organizational change. Lindman received a PhD from the Aalto University School of Economics. Contact him at juho.lindman@hanken.fi.

Tomi Kinnari is a PhD candidate at the Aalto University School of Business and an information security specialist at the Finnish Communications Regulatory Authority's National Cyber Security Centre. His research focuses on open data's business benefits. Kinnari received a MSc in Information and Service Management from Aalto University School of Business in 2012. Contact him at tomi.kinnari@aalto.fi.

Matti Rossi is a professor of information systems at the Aalto University School of Business. His research focuses on open data and enterprise systems. Rossi received a PhD (econ) in Information Systems from University of Jyväskylä on 1998. Rossi won the Technology Academy of Finland's 2013 Millennium Distinction Award for open-source and open-data research. He has been a senior editor of the Journal of the Association for Information Systems (JAIS) and Data Base for Advances in Information Systems, and is currently the editor in chief of Communications of the Association for Information Systems (AIS). He is a member of the IEEE, ACM, and AIS. Contact him at matti.rossi@aalto.fi.

Apps4Finland

Apps4Finland (www.apps4finland.fi) is an annual innovation competition—sponsored by the Finnish Transport Agency, Finnish Transport Safety Agency, Helsinki Regional Transport Authority, and the cities of Tampere, Oulu, and Jyväskylä—that focuses on open-data services and application development.

Between 2009 and 2011, the contest received 193 submissions. From those submissions, we chose 14 companies that offered open-data-related services such as transportation planners, information visualization, and search tools. We then interviewed staff from those companies, including founders, CEOs, project managers, and software engineers. The main article reports some of the interview results.

Abstract

Software specialists know the merits of information visualizations, mashups, and other types of open-data enrichments that serve customers' needs. Commercial services based on these enrichments hold great potential as new businesses. A proposed model categorizes the roles of businesses in enriching open data. This model could help entrepreneurs looking for business opportunities and professionals in companies with underused data resources.

Keywords: data, databases, business computer applications, software development, software engineering

1 INTRODUCTION

Recent research on Semantic Web technologies[1] that can enrich and link datasets for subsequent analysis is creating opportunities for the development of service businesses. This is occurring as governments and private enterprises are increasingly considering the benefits of releasing their datasets openly for use by others.

Governments are opening their datasets to show that they are transparent and open to change[2,3] and to enable service innovation[4]. Many private enterprises also feel that releasing open data can improve their image and even increase profits by using crowds to analyze and utilize the data. Consequently, as organizations have made major data releases, the development of applications and services for enriching and analyzing open data has surged. Businesses that can take advantage of these opportunities could have much to gain. In 2011, the European Commission estimated the economic impacts of aggregate direct and indirect applications and use of public sector information to be in the order of $140 billion annually for the EU27.[5]

From 2012 to 2013, we conducted Next Media Hyperlocal, an industry–academia project studying open datasets and the design of services that exploit them (http://virtual.vtt.fi/virtual/nextmedia/255689/en/read/page.html).[6,7] We gathered and analyzed information from Apps4Finland, an open-data innovation contest (see the sidebar). We interviewed key personnel at 14 companies involved in the contest and examined the firms' roles in the open-data value network. Here, we discuss those roles and their implications for entrepreneurs looking for business opportunities and professionals in companies with underused data resources.

2 OPEN-DATA BUSINESS

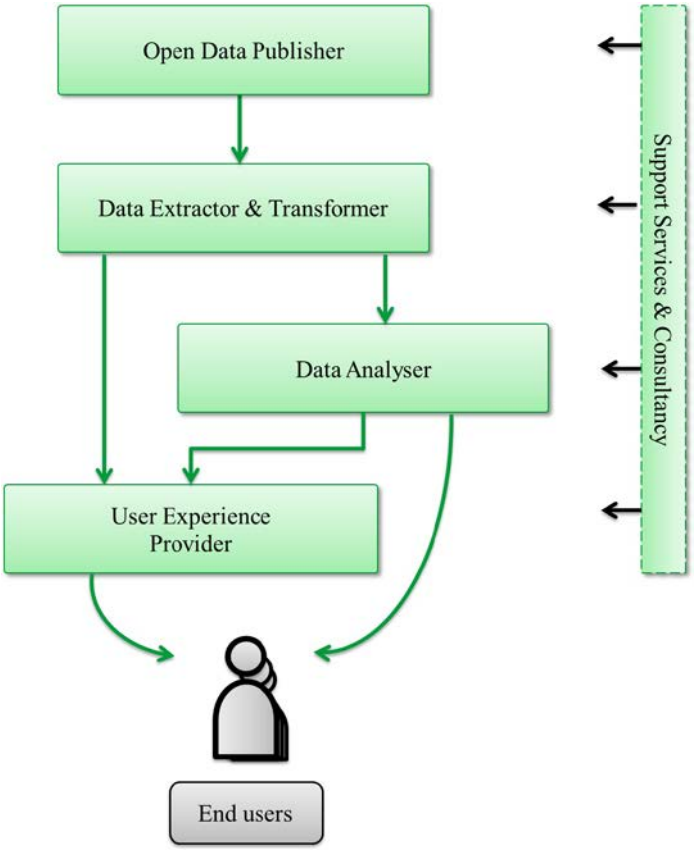
Open Knowledge—a nonprofit group that promotes information openness—says data or content is open if people can freely use, reuse, or redistribute it for any purpose (http://opendefinition.org). The only requirements might be to attribute the source of and share the results obtained with the data.

This makes several commercial service-oriented endeavors possible, even for businesses that don't own any data.

2.1 Roles in the Open-Data Ecosystem

To create sustainable open-data businesses, organizations must understand the open-data value network's five roles. When linked, these roles—whose relationships are shown in Figure 1—form a data-processing chain that enriches raw data into valuable content. The division of work in the resulting open-data community is based on how the information is combined and enriched.[8]

Figure 1 The open-data value network. When linked, these roles form a data-processing chain that enriches raw data into valuable content.



2.2 Open-data publisher

Public organizations such as municipal, state, or federal governments often make data available for use by others. For example, we looked at a local authority (Helsinki Region Transport) that has developed a free API providing access to information that service providers could use to help customers with trip planning. The information included public-transportation routes and timetables, service disruptions, and live data for vehicle location and tracking.

Developers used the API to create about 30 mobile trip-planning applications. Finding data sources on their own and buying the information necessary to provide the services would have been difficult and costly. Meanwhile, the transport authority reaped significant financial gain by opening its data since it didn't have to develop and maintain own mobile clients for different mobile platforms. The public was willing to use, and in many cases also pay for, trip-planning clients developed by third party service providers.

Government agencies could benefit from openly releasing material by demonstrating transparency and providing a public service. Private companies could obtain similar benefits. However, they might be more reluctant to release information because they store a lot of proprietary, employee, and customer data that they want to keep private.

Openly releasing data could be particularly important if it attracts numerous interested parties who could provide many types of useful analyses and services that otherwise would not be available. This has already been done with information the print media has made available.

2.3 Data extractor and transformer

Extraction and transformation entail:

- downloading data;
- converting it to a format suitable for analysis;
- normalizing it to allow cross analysis with other datasets; and removing errors, duplicate information, and fields with no data.

These tasks are time-consuming and require considerable resources. Several companies in our study estimate these tasks consumed at least 50 percent of their total work time.

Nonetheless, many of the companies in our study performed extraction and transformation for themselves because they couldn't find data that was preprocessed for their needs.

We interviewed one group that offers free, community-driven open source tools to ease data extraction and transformation. The community offers free computer scripts that fetch and convert data from several public sources for further processing. Many organizations told that they appreciated the group's work, but that the tools still couldn't find the desired information and thus they had to perform data extraction and transformation themselves.

Commercial opportunities might exist for companies wishing to fill this role. However, offering individual datasets that are diverse enough for multiple customers might be difficult. And providing different datasets for individual customers would be time-consuming and costly.

2.4 Data analyzer

Companies in this role gather and analyze data, often from multiple sources. In our study, these companies were the most successful.

For example, one company in our study uses open data, as well as private data acquired from Finnish firms, to produce credit ratings and other financial information for sale. Another company analyzes business' financial data and draws an easy-to-visualize, tree-shaped image of their balance sheets.

Many analysis services are based on information that is available for long periods of time and kept up to date, such as traffic and weather data. In these cases, it is critical that the information be provided on an ongoing basis in either a standardized format or via API calls.

2.5 User-experience provider

These businesses gather and combine data sources and offer user interfaces or mashups for manipulating the data via a web browser or mobile app.

For example, one company in our study has created a website showing public job-advertisements enriched with additional information from several other sources such as online maps, social media, financial data and news feeds.

User-experience providers' revenue comes either directly from selling mobile apps or website subscriptions, or from advertisements shown within the apps or websites.

2.6 Support-service provider

These companies help the other four value-chain participants with open-data-related tasks. For example, they consult with clients on open-data release procedures, user-experience enhancement, and ways to utilize open data. They might also offer services such as website hosting or data storage. These companies can generate healthy revenue streams if they can maintain high service levels and availability.

3 BUSINESS MODELS

Research has shown that open-data service providers use many business models to create and capture value.[9] Software firms' business models consist of five interrelated elements:[10]

- The offering is the service being provided and involves decisions such as how much customization is desirable and how the product is made available to customers.
- Resources are the assets needed to offer the services to customers and implement the business model.
- Providers must have relationships with other organizations, including customers, partners, and suppliers, to conduct business.
- The revenue model entails factors such as income sources, pricing policies, and cost structure.
- The management mind-set includes issues related to management values, emotions, and attitudes that affect a company's business model.

Table 1 compares the business-model attributes of the roles in the open-data ecosystem.

Table 1 Attributes of roles in the open-data value network.

Role	Offerings	Resources	Relationships	Revenue model	Management mind-set
Open-data publisher	Data for free reuse	Maintaining data storage and the API for developers	Open-data developer community	Cost savings via crowdsourcing (e.g. user interface development)	Constantly improving the API to better serve downstream developers Show the organization as innovative and transparent in the eyes of the public
Data extractor and transformer	Computer scripts that extract, clean, normalize, and prepare data for reuse	Maintaining the code base and documentation	Open-data publishers Open-data developer community	No revenue model	Advance the open data community by providing tools to easier data utilization
Data analyzer	Combine and analyze data from several sources to create visualizations or provide insights	Analyzing data from one or more sources	Open-data developer community, Commercial data sources, Customers	Income from project work	Nobody wants to study spreadsheets. Add value by extracting meaning from raw data and presenting it in simple format
User-experience provider	Create user interfaces for mobile or Web applications	Programming and UI design expertise	Open-data developer community, Commercial data sources, Customers	Advertising Product sales Licensing Freemiums (The basic service is free; extra features require a fee.)	Focus on effectiveness, ease of use, and other usability issues
Support-service provider	Services such as hosting or storage capacity Expertise such as programming or visual design	Programming and UI design expertise	Open-data developer community, Customers	Income from project work	Listen to customer needs and offer solutions that best solve the problems.

4 LOOKING TO THE FUTURE

Several developments are necessary for open-data use to thrive. A basic need is for data owners—particularly those in the public sector—to make their information available to service providers. Public agencies collect valuable information but often don't have the resources to analyze it themselves. They could thus provide the data to analyzers as a service.[11]

Toronto has an interesting open-data policy (http://toronto.ca/open). Its guiding idea is that the city provides all its publicly available data through machine readable formats for easy reuse and service development. Private services built on data provided by the city offer customers many types of valuable information such as restaurant health-inspection results and local beaches' water quality.

The next open-data frontier is private-company data, according to Benjamin Herzberg, the World Bank Institute's program lead for private sector engagement for good governance.[12] He reported on several companies that made their data openly available to provide greater transparency and accountability.

Regardless of the information source, developers will have to continue creating novel services and resources to realize open data's potential.

Using third-party datasets can create managerial and commercial challenges. For example, in Finland maps and lake depth scans are open data, but some of the lake depth scans have been performed by a private company that owns that part of the data and does not allow its use as open data.

Another issue is that many applications that work with open data currently must use labor-intensive data-fetching mechanisms such as screen scraping, comma-separated values, and the analysis of unstructured text and spreadsheets because of the lack of API's and the data being only available in these poorly-structured or non-structured formats.[13,14] Standardized API interfaces and access protocols, such as REST, and data formats could simplify data access and allow for easier reuse of both data and code.[11]

In our research project, we also noticed that some recent open-data services were one-shot or short-term projects by enthusiasts or individual developers. The problem with such services is that, even when users want to keep working with them, they die out

quickly. Experts agree that for the open-data industry to expand, the services must be kept active longer by, for example, making them tenable over greater time periods.

Questions exist about how to ensure datasets' ongoing availability, which will require service providers to develop sustainable revenue models that give them an incentive to keep information up-to-date and accessible. They will also have to come up with unique services to stay in business.

There is a need for middle-tier data-provisioning and -cleansing services to ensure that information is available and in current formats. However, such businesses are not always profitable in the short term because of the small user base. Steps are being taken in this area. For example, the European Commission declared in October 2014 that it would join with Europe's data industry to invest €2.5 billion (about $2.67 billion) to support research and innovation in exploiting big data from both open and private sources (see http://europa.eu/rapid/press-release_IP-14-1129_en.htm).

Our research efforts only scratch the surface of open data's business potential. More research is thus called for, as is further discussion of the important technologies and managerial practices involved.

Specifically, there is a need for survey-based evidence on how best to provide and procure open-data services. One especially interesting research area is the use of both open data and Big Data in fields such as preventive healthcare and traffic planning.

REFERENCES

1. T. Berners-Lee, "Linked Data," *Design Issues,* July 2006, www.w3.org/DesignIssues/LinkedData.html.

2. A. Zuiderwijk and M. Janssen, "Open Data Policies, Their Implementation and Impact: A Comparison Framework," *Government Information Q.*, vol. 31, no. 1, 2014, pp. 17–29.

3. A. Zuiderwijk et al., "Design Principles for Improving the Process of Publishing Open Data," *Transforming Government: People, Process and Policy*, vol. 8, no. 2, 2014, pp. 185–204.

4. G. Kuk and M. Janssen, "Assembling Infrastructures and Business Models for Service Design and Innovation," *Information Systems J.*, vol. 23, no. 5, 2013, pp. 445–469.

5. G. Vickery, *Review of Recent Studies on PSI Re-use and Related Market Developments*, European Commission Directorate General, 2011.

6. J. Lindman, T. Kinnari, and M. Rossi, "Industrial Open Data: Case Studies of Early Open Data Entrepreneurs," *Proc. 47th Hawaii Int'l Conf. System Sciences* (HICSS 14), 2014, pp. 739–748.

7. J. Lindman, M. Rossi, and V.K. Tuunainen, "Open Data Services: Research Agenda," *Proc. 46th Hawaii Int'l Conf. System Sciences* (HICSS 13), 2013, pp. 1239–1246.

8. A. Latif et al., "The Linked Data Value Chain: A Lightweight Model for Business Engineers," *Proc. 5th Int'l Conf. Semantic Systems* (I-Semantics 09), 2009, pp. 568–577.

9. A. Osterwalder, "The Business Model Ontology: A Proposition in a Design Science Approach," PhD thesis, Faculty of Business and Economics, Univ. of Lausanne, 2004.

10. R. Rajala, "Determinants of Business Model Performance in Software Firms," PhD dissertation, Helsinki School of Economics, 2009.

11. N. Huijboom and T. Van den Broek, "Open Data: An International Comparison of Strategies," *European J. ePractice*, vol. 12, no. 1, 2011, pp. 1–13.

12. B. Herzberg, "The Next Frontier for Open Data: An Open Private Sector," blog, World Bank, 3 Mar. 2014; http://blogs.worldbank.org/voices/next-frontier-open-data-open-private-sector.

13. C. Bizer, T. Heath, and T. Berners-Lee, “Linked Data—the Story So Far,” *Int'l J. Semantic Web and Information Systems*, vol. 5, no. 3, 2009, pp. 1–22.

14. T. van den Broek, A. F. van Veenstra, and E Folmer, “Walking the Extra Byte: A Lifecycle Model for Linked Open Data”, In: E. Folmer, M Reuvers, and W Quak, Pilot Linked Open Data Nederland, Deel 2 – De Verdieping, Linked Open Data, Amersfoort, 2013, pp. 94 – 110. http://doc.utwente.nl/93940/1/6.boek2.pdf.